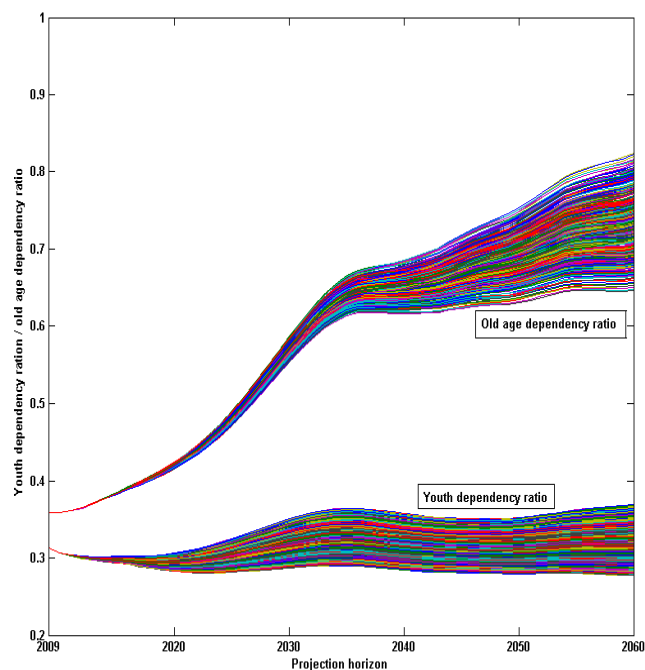




Assessing the uncertainty in population projections

A test based on the 12th coordinated population projection for Germany

Dirk Zeitz



EUR 24355 EN - 2010

The mission of the JRC-IPSC is to provide research results and to support EU policy-makers in their effort towards global security and towards protection of European citizens from accidents, deliberate attacks, fraud and illegal actions against EU policies.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information

Address: Via E. Fermi 2749, 21027 Ispra, Italy
E-mail: Dirk.Zeitz@jrc.ec.europa.eu
Tel.: +39 0332 786033
Fax: +39 0332 785733

<http://ipsc.jrc.ec.europa.eu/>
<http://www.jrc.ec.europa.eu/>

Legal Notice

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

***Europe Direct is a service to help you find answers
to your questions about the European Union***

**Freephone number (*):
00 800 6 7 8 9 10 11**

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server <http://europa.eu/>

JRC 58249

EUR 24355 EN
ISBN 978-92-79-15709-7
ISSN 1018-5593
DOI 10.2788/90191

Luxembourg: Publications Office of the European Union

© European Union, 2010

Reproduction is authorised provided the source is acknowledged

Printed in Italy

Assessing the uncertainty in population projections

A test based on the 12th coordinated population projection for Germany

By Dirk Zeitz*

Abstract:

In this paper I use variance based sensitivity analysis to investigate the uncertainty in projections of future population arising from the assumptions taken. Instead of for a limited number of scenarios the uncertainty is explored within the range of assumptions for the model parameters through sampling.

To this purpose, the cohort component method and the assumptions taken in 12th coordinated population projection for Germany have been reengineered in a sensitivity analysis setting. This allows attributing the uncertainty in the outputs of the demographic model (e.g. size of the population) to the uncertainty in the model inputs (e.g. total fertility rates).

Keywords:

Demographics, Cohort-component method, Population projections, Uncertainty analysis, Variance based sensitivity analysis.

* European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen, Via E. Fermi 2749, 21027 Ispira, Italy. Tel.: +39 0332 786033; E-mail: Dirk.Zeitz@jrc.ec.europa.eu

TABLE of CONTENTS

1	Introduction and motivation.....	3
2	The cohort component method employed for long-term population projections ..	4
2.1	Basic statement of the cohort-component model.....	4
2.2	Accounting for uncertainty in population projections	6
3	Variance based sensitivity analysis.....	7
3.1	Theoretical background	7
3.2	Sample generation.....	8
4	Implementation of the case study	9
4.1	Assumptions made in the 12 th coordinated population projections	9
4.1.1	Fertility rates	9
4.1.2	Migration.....	10
4.1.3	Life expectancy	11
4.2	The numerical implementation	13
5	Discussion of results	15
5.1	Visual sensitivity analysis with scatter-plots	16
5.2	Estimated sensitivity indices.....	19
6	Conclusions.....	23
7	Acknowledgements.....	24
	References.....	25

LIST of FIGURES

Figure 1	Age structure of mothers (target values)	10
Figure 2	Mortality for males (life table 2006/2008 and scenarios)	13
Figure 3	Sampling ranges for total fertility rates over the projection horizon.....	14
Figure 4	Population projected for 2060	15
Figure 5	Youth dependency and old age dependency ratio	16
Figure 6	Scatter-plot Size of population vs. model inputs (T=2009)	17
Figure 7	Scatter-plot Size of population vs. model inputs (T=2060)	17
Figure 8	Scatter-plot Youth dependency ratio vs. model inputs (T=2060)	18
Figure 9	Scatter-plot Old age dependency ratio vs. model inputs (T=2060).....	19
Figure 10	Sensitivity indices: Size of population	20
Figure 11	Sensitivity indices: Youth dependency ratio	21
Figure 12	Sensitivity indices: Old age dependency ratio.....	21

1 Introduction and motivation

Prophecy about the future of societies is an uncertain business. However, the development of future population is a hot issue for policy makers as it requires decisions to be taken today.

An ageing population has important implications on pension systems and age related costs. Consequently a number of States have designed policies to address this issue and the resulting demographic changed. Measures include for instance a raise of the retirement age, or shortening periods of education in order to reduce the age of people at which they enter labour markets. In Europe with a few exceptions (like for instance France) total fertility rates are below the natural replacement rate. Since fewer children are born, the size of the potential future workforce is to decline. This is a relevant issue since social security systems (in particular pension systems) depend to a large extent on the ratio of persons contributing to the system and those receiving from the schemes. If this ratio is to change in future, also these systems need to be adopted accordingly in order to ensure fiscal sustainability.

As means to address the information needs for designing social security systems policy makers rely on population projections. Population projections have a long standing history. In terms of mathematical models one refers to periodic matrix models. However, as the time horizon is extended the numbers offered yields less and less information. Not surprisingly, most explorations of demographic futures profess to offer not predictions of actual developments but “projections” – calculations that work out the numerical implications of well-specified plausible or illustrative assumptions about the behaviour of the factors that determine demographic change (see Worldbank, United Nations, Eurostat 2006, U.S. Census Bureau). While we can predict the population development within the range of assumptions taken, so far it has not been investigated how the model inputs drive the uncertainty in the population projections. This is the more interesting since tools are available that allow modellers to quantify the uncertainty in the model output attributed to its model inputs.

In this paper I investigate the uncertainty arising from the assumptions taken in population projections. Instead of a few scenarios a high number of samples is used and variance based sensitivity analysis is employed in order to quantify the uncertainty in the projected size of the population and in some further measures. The uncertainty in the model outputs is in a further step attributed to the uncertainty in model inputs. For the case study I have reengineered the cohort component method in a sensitivity analysis setting and assess the assumptions made in the 12th coordinated population projection for Germany that was presented in November 2009 (See Statistisches Bundesamt 2009).

The paper is organised the following way. In section 2, I provide a brief description of the standard model used for long-term population projections – the cohort component method. Section 3 gives an overview about the applied technique of the sensitivity analysis. In section 4 the numerical implementation of the cohort-method in a sensitivity analysis setting is outlined. Section 5 contains the results of the numerical experiments and I conclude in section 6.

2 The cohort component method employed for long-term population projections

Long-term population projections employ the cohort component method. In this model initial populations for countries or regions are grouped into cohorts defined by age and sex, and the projection proceeds by updating the population of each age- and sex-specific group according to assumptions about three components of population change: fertility, mortality, and migration. Each cohort survives forward to the next age group according to assumed age-specific mortality rates.

The cohort component method was formalised in mathematical terms by Leslie (1945), and first employed in producing a global population projection by Notestein (1945).¹ Since Notestein's 1945 projection, the cohort-component method has become the dominant means of projecting population and has remained essentially unchanged, except for extensions to multi-state projections and innovations in characterizing uncertainty.

A fundamental feature of the method is that the projected size and age structure of the population at any point in the future depends entirely on the size and age structure at the beginning of the period and the age-specific fertility, mortality, and migration rates over the projection period. Uncertainty in projection outcomes arises not from uncertainty in the formal projection model itself, but from uncertainty in the baseline population data and the assumptions of future trends in vital rates.

2.1 Basic statement of the cohort-component model

The cohort-component model links demographic processes and changes in population characteristics. The cohort-component method for estimating and projecting a population is distinguished by its ability to preserve knowledge of an age distribution of a population over time. It is a special case of the component model, which is defined by the use of estimates or projections of births, deaths, and net migration to update population. In its simplest statement, the component method is expressed by the following equation:

$$N_t = N_{t-1} + B_{t-1,t} - D_{t-1,t} + M_{t-1,t}.$$

In the equation N_t and N_{t-1} denote the population at time t ; and at time $t-1$. The variables $B_{t-1,t}$, $D_{t-1,t}$ and $M_{t-1,t}$ denote the births (B), deaths (D) and the net migration (M) in the interval from time $t-1$ to time t .

The cohort component method is based on a similar logic for individual age groups, recognising that the source population for a given age group is the population at time $t-1$ in adjacent younger age group. For the initial age group, it is births during the interval from $t-1$ to t . The equation is replaced by two equations, depending on whether the age group is zero (meaning under 1) or any other age as of the last birthday denoted by k .

¹ The first population projection based on the component method was made, however, half a century earlier (Cannan (1895) for UK and Wales).

$$N_{0,t} = B_{t-1,t} - D_{0,t-1,t} + M_{0,t-1,t}$$

$$N_{k,t} = N_{k-1,t-1} - D_{k,t-1,t} + M_{k,t-1,t}$$

In matrix formulation including the mathematical operations the cohort component method can be described by:

Matrix I: female population at the end of year t+1

$$\begin{pmatrix} N_{0,t+1}^F \\ N_{1,t+1}^F \\ N_{2,t+1}^F \\ \vdots \\ \vdots \\ N_{k,t+1}^F \end{pmatrix} = \begin{pmatrix} 0 & \dots & sF_{15,t}P_{-1,t}^F & \dots & sF_{49,t}P_{-1,t}^F & 0 & \dots & 0 \\ p_{0,t}^F & 0 & 0 & & \dots & \dots & 0 & \dots & 0 \\ 0 & p_{1,t}^F & 0 & & \dots & \dots & 0 & \dots & 0 \\ \vdots & 0 & \ddots & & \ddots & & & & \\ \vdots & \vdots & & & 0 & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & & \dots & \dots & 0 & p_{k-1,t}^F & 0 \end{pmatrix} \bullet \begin{pmatrix} N_{0,t}^F \\ N_{1,t}^F \\ N_{2,t}^F \\ \vdots \\ \vdots \\ N_{k,t}^F \end{pmatrix} + \begin{pmatrix} M_{0,t}^F \\ M_{1,t}^F \\ M_{2,t}^F \\ \vdots \\ \vdots \\ M_{k,t}^F \end{pmatrix}$$

Matrix II: male population at the end of year t+1

$$\begin{pmatrix} N_{0,t+1}^M \\ N_{1,t+1}^M \\ N_{2,t+1}^M \\ \vdots \\ \vdots \\ N_{k,t+1}^M \end{pmatrix} = \begin{pmatrix} 0 & \dots & (1-s)F_{15,t}P_{-1,t}^M & \dots & (1-s)F_{49,t}P_{-1,t}^M & 0 & \dots & 0 \\ p_{0,t}^M & 0 & 0 & & \dots & \dots & 0 & \dots & 0 \\ 0 & p_{1,t}^M & 0 & & \dots & \dots & 0 & \dots & 0 \\ \vdots & 0 & \ddots & & \ddots & & & & \\ \vdots & \vdots & & & 0 & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & & \dots & \dots & 0 & p_{k-1,t}^M & 0 \end{pmatrix} \bullet \begin{pmatrix} N_{0,t}^M \\ N_{1,t}^M \\ N_{2,t}^M \\ \vdots \\ \vdots \\ N_{k,t}^M \end{pmatrix} + \begin{pmatrix} M_{0,t}^M \\ M_{1,t}^M \\ M_{2,t}^M \\ \vdots \\ \vdots \\ M_{k,t}^M \end{pmatrix}$$

where,

- $N_{x,t}^M$ respectively $N_{x,t}^F$ as male respectively female population of x-year old age cohort at the end of period t .
- $F_{x,t}$ as the age specific birth rate of the x-year old mother in year t . This corresponds to the age-specific birth rate of one woman.
- $p_{x,t}^M$ respectively $p_{x,t}^F$ as the probability of x-year old woman or man to reach the next year of age.
- $M_{x,t}^M$ and $M_{x,t}^F$ as sum of the female or male migrants of age cohort x in year t .
- s as the relation of female to male newborns (often 1:1.0563 in simulations);
- $x=0,1,2,\dots,k-1,k$ where k denotes the oldest age-cohort.

New born females and males are determined by

$$N_{0,t+1}^F = s \cdot p_{-1,t}^F \cdot \sum_{x=15}^{49} F_{x,t} \cdot N_{x,t}^F ,$$

and

$$N_{0,t+1}^M = (1-s) \cdot p_{-1,t}^M \cdot \sum_{x=15}^{49} F_{x,t} \cdot N_{x,t}^F$$

Thus, births are calculated by age-specific birth rates ($F_{x,t}$) multiplied by the number of women within specific groups of cohorts ($N_{x,t}$), usually of age between 15 and 49.

2.2 Accounting for uncertainty in population projections

The fact that projecting fertility, mortality and migration plays a central role in the cohort-component method allows demographers to draw on specialized knowledge of each of these components of population change. Institutions therefore normally project trends in vital rates based on expert opinion (see O'Neill et al. (2002)). Historically, however, it has been difficult to determine precisely how knowledge has been applied to such projections. Ahlburg and Lutz (1998) argued that assumptions and reasoning often have been hidden behind a “veil of secrecy”.

The uncertainty associated with a population projection is dealt with mainly in two ways (see Keilman, et al. 2002): scenarios and probabilistic projections.

The most common approach is to present alternative scenarios (as in the case of 12th coordinated population projection) that assume higher or lower vital rates than in the medium or central scenario. The strength of the alternative scenario approach is that in many cases users may need alternatives to a single central scenario, but still require self-consistent, independent scenarios rather than confidence intervals around a most likely projection. However, the approach also has several weaknesses. The most important is that if no specific level of uncertainty is associated with the alternatives, it is not possible for users to interpret the precise meaning of the ranges presented. Another problem with the scenario approach is that the choice of certain values for some assumptions may mean that choices for others are unreasonable.

An alternative to scenarios as a means of communicating uncertainty is to explicitly account for uncertainty in projected trends of fertility, mortality, and migration, and derive the resulting probability distributions for projected population size and age structure. There have been three main bases for determining the probabilities associated with vital rates: expert opinion, statistical analysis, and analysis of errors in past projections. However, there are a number of obstacles to this approach, including deciding who constitutes an expert and counteracting the observed conservatism in the projection of future trends.

Statistical analysis of historical time series data can be used either to project population size directly or to generate probability distributions for population size or vital rates. Lee (1998) argues that, unlike methods based purely on expert opinion, these methods are capable of producing internally consistent probability distributions. The analysis of errors in historical projections can be used as a basis for generating uncertainty intervals around a projection produced by some other means. This method relies on the assumption that current projections are subject to errors similar to those made in the past, although trends such as improvements in the quality of data on the

initial population at the start of the projection can be controlled for. Its strength is that it yields a probability distribution for a given projection that is consistent with the essential features of errors observed in the past (NRC 2000).

In the investigation here, we explore the space of uncertainty in the population projection by using multiple model evaluations. Instead of considering only a limited number of scenarios we take a number of randomly sampled points within the range of the assumptions taken for the parameters (for instance between the “High” and “Low” scenarios for total fertility rates), each one reflecting an independent scenario. This allows us to employ variance based sensitivity analysis and to attribute the uncertainty in the outputs of the model (size of population in t , age-dependency ratios) to the model inputs.

3 Variance based sensitivity analysis

Modern statistical tools such as sensitivity analysis are means to assess the impact of the assumptions taken about the factors determining demographic change. Recently, variance based methods have become very popular among practitioners of global sensitivity analysis. Variance based methods have a long history in sensitivity analysis, starting with a Fourier implementation in the seventies (Cukier et al., (1973)) and having a milestone in the work of Sobol’ (1993), while total sensitivity indices were introduced by Homma and Saltelli (1996).

3.1 Theoretical background

Here, just an overview of variance decomposition methods is provided. For further details, see Sobol’ (1993). Given a square integrable function $Y = f(X_1, X_2, \dots, X_k)$ defined over Ω , the k -dimensional unit hypercube:

$$\Omega = (X \mid 0 \leq x_i \leq 1; i = 1, \dots, k),$$

the decomposition of the model output variance, which holds for independent model inputs, X is:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \dots + V_{12\dots k} \quad (1)$$

where

$$V_i = V(f_i(X_i)) = V_{X_i}(E_{X_{\sim i}}(Y \mid X_i)) \quad (2)$$

$$V_{ij} = V(f_{ij}(X_i, X_j)) = V_{X_i X_j}(E_{X_{\sim ij}}(Y \mid X_i, X_j)) - V_{X_i}(E_{X_{\sim i}}(Y \mid X_i)) - V_{X_j}(E_{X_{\sim j}}(Y \mid X_j))$$

and so on for higher order terms. If we divide both sides of (1) by $V(Y)$, we obtain:

$$\sum_s S_s + \sum_i \sum_{j>i} S_{ij} + \dots + S_{12\dots k} = 1.$$

Since the number of terms in the decomposition of the variance of a model with k model inputs grows at 2^k , in sensitivity analysis it is customary to compute just two

sets of k indices: the k so-called ‘first order’ effects (S_i) and the k so-called ‘total’ effects (S_{Ti}):

$$S_i = \frac{V_{X_i}(E_{X_{\sim i}}(Y | X_i))}{V(Y)} \quad (3)$$

$$S_{Ti} = \frac{E_{X_{\sim i}}(V_{X_i}(Y | X_{\sim i}))}{V(Y)} = 1 - \frac{V_{X_{\sim i}}(E_{X_i}(Y | X_i))}{V(Y)}. \quad (4)$$

Details on total effects are offered by Homma and Saltelli (1996). In order to obtain first order and total sensitivity estimates samples are generated to subsequently evaluate the model.

3.2 Sample generation

In order to evaluate the model we generate a sample of N points within Ω , the input space of dimension k . The coordinates of the N points can be summarised by a matrix A with N rows and k columns. Let us generate a second set of N points within Ω , independently from the previous set, and summarise the coordinates in a matrix B . Now we introduce a third set of matrices, $A_B^{(i)}$ for $i=1 \dots k$ where all columns are from A except the i -th column which is from B .² Hence, the matrix for the evaluation of the model is made from 6 matrices (A , B and k times $A_B^{(i)}$). We consider the first order estimator as provided by Saltelli/Sobol’:

$$V_{X_i}(E_{X_{\sim i}}(Y | X_i)) = \frac{1}{N} \sum_{j=1}^N f(B)_j (f(A_B^{(i)})_j - f(A)_j) \quad (5)$$

and for total order effects the one offered by Jansen (1999). This estimator for S_{Ti} proceeds via $E_{X_{\sim i}}(V_{X_i}(Y | X_{\sim i}))$ rather than $V_{X_{\sim i}}(E_{X_i}(Y | X_{\sim i}))$:

$$E_{X_{\sim i}}(V_{X_i}(Y | X_{\sim i})) = \frac{1}{2N} \sum_{j=1}^N (f(A)_j - f(A_B^{(i)})_j)^2 \quad (6)$$

The estimators for first order indices S_i (Saltelli/Sobol’) and for total effects indices S_{Ti} (Jansen) are considered as the most efficient for the estimation of variance-based sensitivity indices (see Saltelli et al. (2010)). Thus, in order to compute both sets of S_i and S_{Ti} for the k inputs just the triplet of matrices A , B , $A_B^{(i)}$ is needed. Finally, dividing formulas (5) and (6) by $V(Y)$ leads to the first order and total effects indices, respectively. Note that the total output variance $V(Y)$ is estimated using only the outputs $f(A)_j$ and $f(B)_j$, as they are independent, while the outputs $f(A_B^{(i)})_j$ are excluded from the computation as they are not independent.

² The model evaluation matrix of size $(k+2)*N \times k$ has uncorrelated columns.

4 Implementation of the case study

4.1 Assumptions made in the 12th coordinated population projections

The case study to test the uncertainty of the model is the 12th coordinated population projection for Germany. Data have been taken as provided on the web of the Federal Statistical Office. Furthermore, the Federal Statistical Office provided some more detailed information on the assumptions taken.

4.1.1 Fertility rates

Three different assumptions are taken with respect to the number of children per woman. This measure of the ‘so-called’ total fertility rate (for instance 1.9 children per woman) is the sum over the age-specific fertility rates ($F_{x,t}$) of women in their child-bearing years (assumed to be in the age interval 15-49).

Box 1 Assumptions on birth rates

Scenario	Children per woman	Average age of birth
Baseline (2006-2008)	1.36 children per woman	29.8 years
F1	2009-2060 1.4 children per woman	Raise to 31.4 until 2020, then constant
F2	Raise to 1.6 until 2025; 2026-2060 1.6 Children per woman	Raise to 30.9 until 2025 then constant
F3	Decrease to 1.2 children per woman until 2060	Raise to 31.9 until 2060

The three scenarios for the total fertility do not only affect the number of children per woman, but they provide at the same time for different average age structures of mothers at birth (see Figure 1). The scenarios are based on different trends which have been observed over the medium- and long-term trend. In more detail:

- a) F1 foresees a continuation of the most important long-term trends until 2020:
- More and more women would become mothers only after their 30th birthday. The frequency of births of younger mothers would further reduce,
 - the share of mothers with three and more children would slightly decrease,
 - the share of women without children would slightly increase and would then remain constant.

Under these assumptions the total fertility rates remains constant at a level of 1.4 children per woman. At the same time the average age of giving birth raises by 1.6 years.

- b) F2 foresees a trend reversal in the medium term which leads to an improvement of the recent situation:

- The frequency of births of under 30-years old would stabilize,
- the 30-years and older women would have more children than today,
- the births postponed to a higher age would take place in fact, such that childlessness would not increase further.

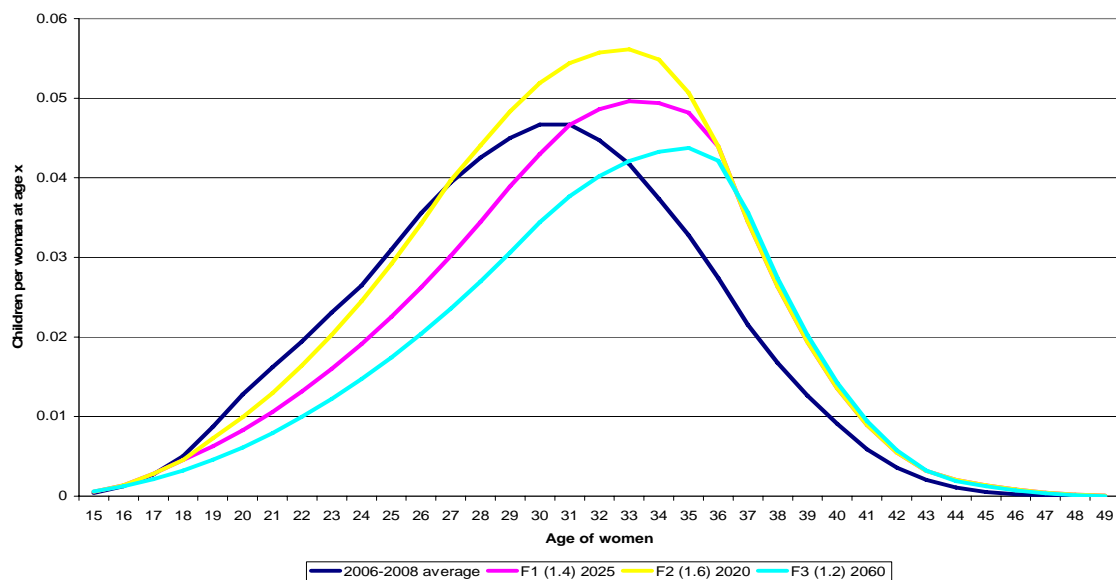
In such a scenario the total fertility rate would increase to 1.6 children per woman by 2025. The average age of giving birth would at the time raise by 1.1 years.

c) F3 the observed trends would be extrapolated. The childlessness would approach a level not yet seen:

- More and more women would become mothers after their 30th birthday, the frequency of birth of younger women would decrease,
- the share of mothers with three and more children strongly decreases as women become mothers later,
- the share of women without children continuously increases as the initially postponed births would be realised to a lower extent.

In such scenario the total fertility rate would decrease to 1.2 children per woman until 2060. The average age at giving births would rise by 2.0 years at the same time.

Figure 1 Age structure of mothers (target values)



4.1.2 Migration

Regarding migration, the scenarios provide for different assumptions on the *net migration surplus*.

Box 2 Assumptions on annual migration surplus

Year	M1	M2
2009	-30000	-30000
2010	10000	10000
2011	40000	40000
2012	60000	80000
2013	80000	100000
2014	100000	120000
2015	100000	140000
2016	100000	160000
2017	100000	170000
2018	100000	180000

2019	100000	190000
2020-2060	100000	200000

The scenarios foresee from 2020 an annual net migration surplus of 100.000 (M1), or 200.000 migrants (M2). For the period 2009-2020, starting from a negative net migration surplus in 2009, both scenarios consider a stepwise increase in the number of migrants. The two scenarios do not only differ in the absolute number of the net migration surplus but they also have different underlying assumptions regarding the structure of migrants in terms of age and gender. The Federal Statistical Office has made the following assumptions as regards the age cohorts of migrants.³

Box 3 Assumptions on the age cohorts of migrants (from 2020)

	M1		M2	
	male	female	male	female
under 15	0.1163	0.0874	0.0840	0.0849
15-20	0.1199	0.1440	0.0757	0.1131
20-25	0.4029	0.4117	0.2713	0.3524
25-30	0.2257	0.1736	0.2036	0.1789
30-35	0.0890	0.0797	0.1247	0.0959
35-40	0.0324	0.0371	0.0814	0.0551
40-45	0.0337	0.0310	0.0677	0.0421
45-50	0.0441	0.0368	0.0636	0.0405
50-55	0.0305	0.0277	0.0425	0.0299
55-60	0.0090	0.0093	0.0175	0.0144
60+	-0.1034	-0.0383	-0.0322	-0.0072
	1.0000	1.0000	1.0000	1.0000

A third difference between the scenarios relates to the ratio between female and male migrants. In migration scenario M1 the male-female ratio is 45:55 (for the projection horizon from 2014) whereas it is 105:95 in migration scenario M2 (from 2020).

4.1.3 Life expectancy

Further assumptions concern life expectancies for females and males over the projection horizon. Life expectancy relates to the expected years yet to live on at a given age. This measure is obtained from mathematical operations (the probability to survive $p_{x,t}$) in the so-called life table. These operations involve the following:

Assume an initial population $I(0)$ at time 0 . In time $t+1$ the population is $I(1)=I(0)(p_x)$. The total years lived of the entire population at the end of the first period is $L(0)=I(0)(p_x+a(x)d(x))$, where $a(x)$ is a factor for the fraction of the years that the ones that died during the particular year contributed to the total lived through years of the entire population ($L(x)$).⁴

³ The assumptions on migrants do not foresee ageing migrants. The age structure corresponds to the mean value for the immigrating and emigrating people years 2005-2007.

⁴ In the year 0 it is $a=0.15$ as in case of the sudden child death most newborn die soon after birth. For later periods it is assumed $a=0.5$, i.e. people died uniformly distributed over the year.

The operations involve all age cohorts. The sum over $L(x)=e_x l_x$ gives the total yet to live through years of the initial population. The sum of $L(x)$ over all age cohorts divided by the survivals at the beginning of the period ($l(x)$) gives the life expectancy at a given age. The life expectancy at a given age is the sum of all yet to live through years of all age cohorts with a completed age ($x+i$) higher than the actual age cohort (x) divided by the number of survivals at the beginning of the period (x). Box 4 below provides an indicative description of the life table operations.

Box 4 Life table males, Germany 2006/2008

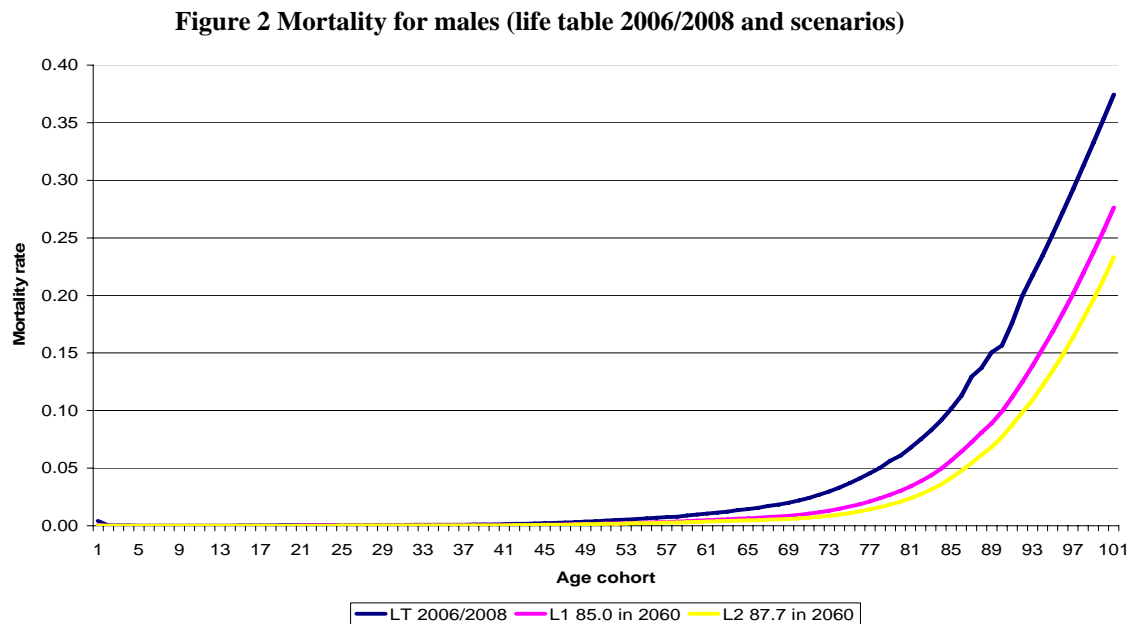
completed age	Probability of		Survivals at age x	Deceased at age x till x+1	Of survivals at age x		Average life expectancy at age x in years
	death from age x to x+1	survival			until age x+1 lived through	total yet to live through years	
x	d_x	p_x	l_x	d_x	L_x	$e_x l_x$	e_x
0	0.004129	0.995871	100 000	413	99 650	7 716 667	77.17
1	0.000341	0.999659	99 587	34	99 570	7 617 016	76.49
...
20	0.000607	0.999393	99 196	60	99 166	5 727 751	57.74
...
40	0.001424	0.998576	97 732	139	97 663	3 756 614	38.44
...
100	0.374448	0.625552	698	261	567	1 382	1.98

The scenarios developed for life expectancy are based on the short-term and long-term trends which were observed in the past. For each age cohort a short-term trend since 1970 and a long-term trend since 1871 in the mortality rates was calculated. In particular, the short-term trend (L2) takes into account the considerable decrease in mortality of the age cohorts from 60 years of age on which was observed in the last 35 years. The scenario L1 is a combination of the short-term trend development since 1970 and the long-term trend development since 1871.

Box 5 Assumptions on life expectancy

Life expectancy at birth			
	2006/2008	2060 L1	2060 L2
Men	77,2	85,0	87,7
Women	82,4	89,2	91,2
Life expectancy at the age of 65			
	2006/2008	2060 L1	2060 L2
Men	17,1	22,3	24,7
Women	20,4	25,5	27,4

The dates for the life expectancy correspond to the mortality rates as presented in Figure 3. The curves L1 and L2 are the values that the Federal Statistical Office presumed for the mortality for all age-cohorts in hypothetical life tables for the year 2060.



4.2 The numerical implementation

In the projection of the Federal Statistical Office twelve scenarios have been calculated.⁵ In the present study 500 model evaluations are carried out. The challenge was to implement the model and the assumption taken in a sensitivity analysis setting. For each year of the projection horizon a sensitivity analysis was made which enables the exploration of the sensitivity indices for the model inputs for each year of the projection.

The data on the stock of the population (as of 31 December 2008) are taken from the publicly available information on the web page of the Statistical Office.⁶ Detailed information on age-specific birth rates, migration and the life tables used in the projection was provided by the Statistical Office on request.

We have reengineered the cohort-component method and the assumptions made in the 12th coordinated population projection. However, there remain some differences with respect to the assumptions made in the projection.

Simplified assumptions for the SA:

For each model input (total fertility rate, life expectancy women, life expectancy men, and migration) we generate a random sample of size N in the interval 0 to 1 from a uniform distribution.⁷ The uniform distribution was chosen as there was no indication why different probabilities should be allocated to a particular scenario (for instance a

⁵ The 12th population projection provided three scenarios for fertility rates, two for the net migration surplus and a further two for life expectancy (females and males).

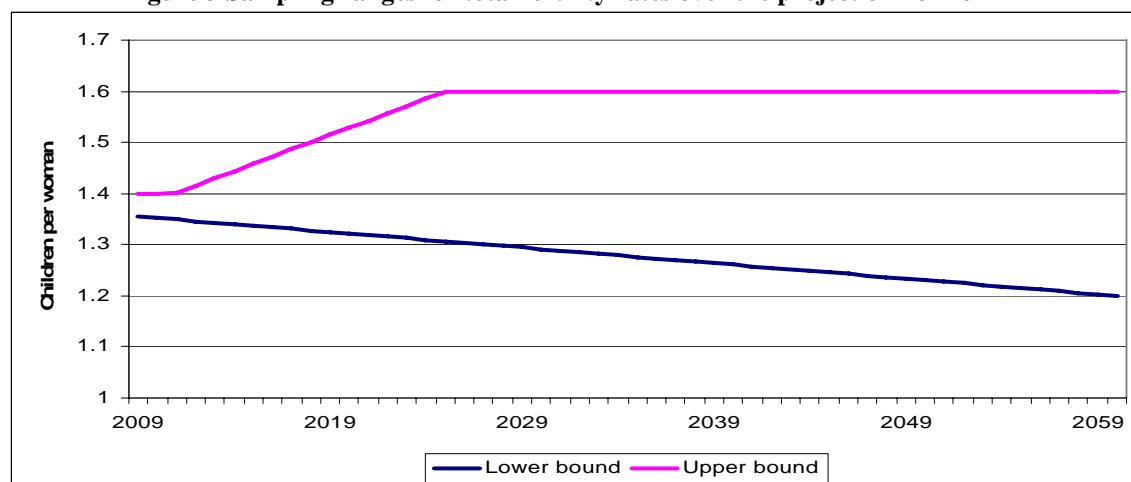
⁶ Data on the stock of population as of 31.12.2008 was taken from Fachserie 1 Reihe 1.3 – 2008.

⁷ The sample from a uniform distribution can relatively well cover the entire space of uncertainty in the range of the assumptions.

mean scenario for total fertility rate). Moreover, the uniform distribution ensures that the uncertainty within the range of assumptions is well captured in the model.

The extreme values of the sample range reflect the bounds of the assumptions made for the four model inputs. For instance a value of 0 corresponds to a total fertility rate of 1.2 (F3), a value of 1 means a total fertility rate of 1.6 (F2). The full range of uncertainty in the total fertility rate (1.2 to 1.6 children per women) is the largest only for the projection for the year 2060 (reflecting the path of the trajectories over time as shown in Figure 3 below).

Figure 3 Sampling ranges for total fertility rates over the projection horizon



The approach takes into account the time dependency of the total fertility rate (as provided for in the scenarios) over the projection horizon. Furthermore, not only the ranges of the *total fertility rates* are different in the three scenarios, but also the age specific birth rates of mothers. Accordingly, a sample value of 1 implies apart from a high total fertility rate also younger mothers, while values near zeros would not only imply a lower child-bearing rate but also relatively “old” mothers.⁸

There are further differences between the reengineered assumptions and those made in the coordinated population projection with regard to the *net migration surplus*. Whereas the two scenarios (M1 and M2) differ in three factors – size of net migration surplus, age-structure of migrants and gender relation, we treat only the absolute number of migrants as uncertain. The age structure and the gender relation of migrants are held fixed throughout the projection. The values for those parameters are basically a mean scenario of the two scenarios. The ratio between female and male migrants for the total number of migrants was set to 50:50. The sample provides N values between the lower bound and the upper bound for the net migration surplus (see Box 2).

Regarding life expectancies for females and males separate random samples were generated. Starting from the data provided in the life table 2006/2008 for women and men, the corresponding mortality rates approach the range of values specified in the

⁸ Please note that the age of mothers was not treated as a uncertain model input (independently of the total fertility rate). For future work we are planning to explore a richer model that may include further uncertain parameters of the model.

life tables for the year 2060. For each year of the projection horizon the mortality rate for each age cohort is reduced by a percentage to reach the respective values (for L1 and L2) of the year 2060. A sampled value of zero for the life expectancy corresponds to the scenario L1 (moderate increase in life expectancies, L1). Accordingly, a value of 1 would reflect a strong increase in life expectancy (L2).

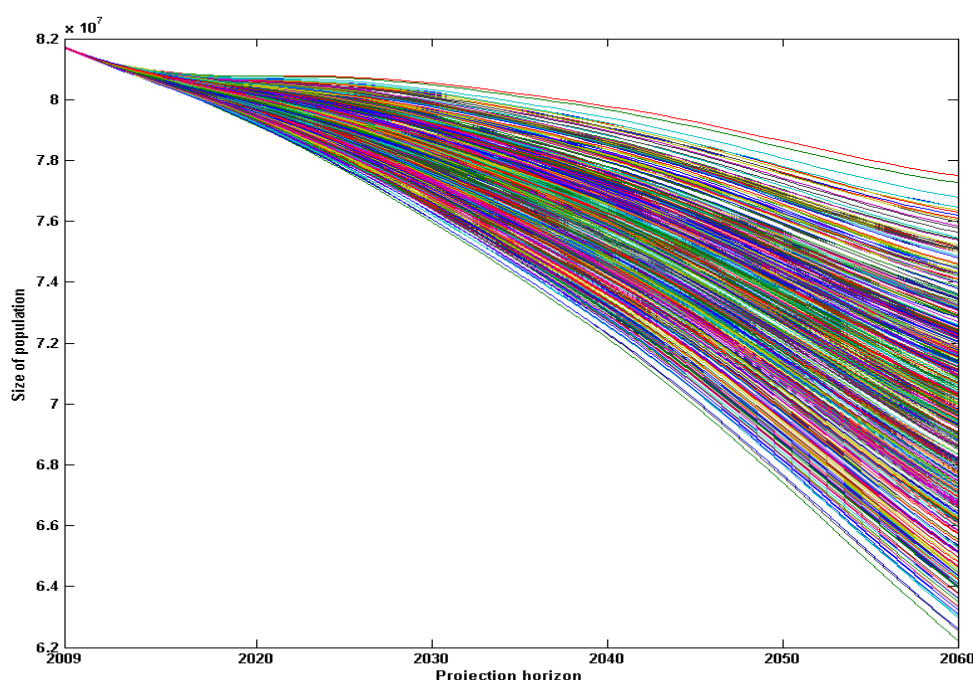
Model outputs

The variance based sensitivity analysis focused on three model outputs. Besides the size of population, two more model outputs have been investigated – the old age dependency ratio and the youth dependency ratio. From a policy maker's perspective the latter two output measures are the probably the more interesting ones. They measure the ratio of the number of persons younger than 20 years and over 64 years of age, respectively, as of the number of persons in working age, i.e. those persons between 20 and 64 years.⁹

5 Discussion of results

For 500 model evaluations the mean projected mean population in 2060 would be 69,451 million. The minimum population in the projection is 61,988 million, the maximum population 77,748 million.

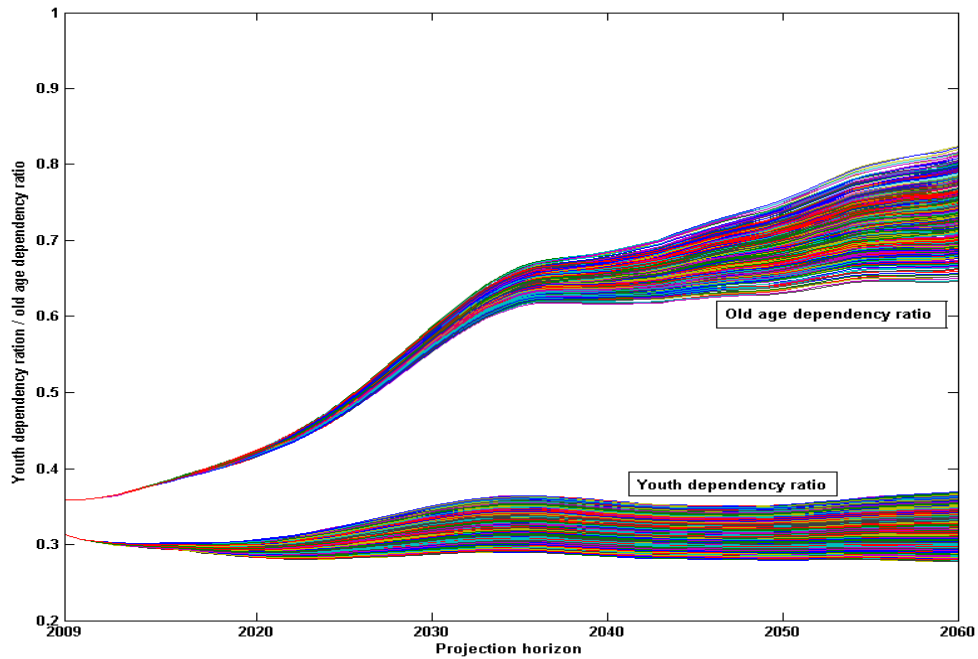
Figure 4 Population projected for 2060



Not only the size of population is subject to change over the projection horizon but so is the age structure of the population. This is presented in Figure 4 which shows the developments of the youth dependency ratio and the old-age dependency ratio.

⁹ In most studies the working age population encompasses those persons between 20 years to 64 years of age.

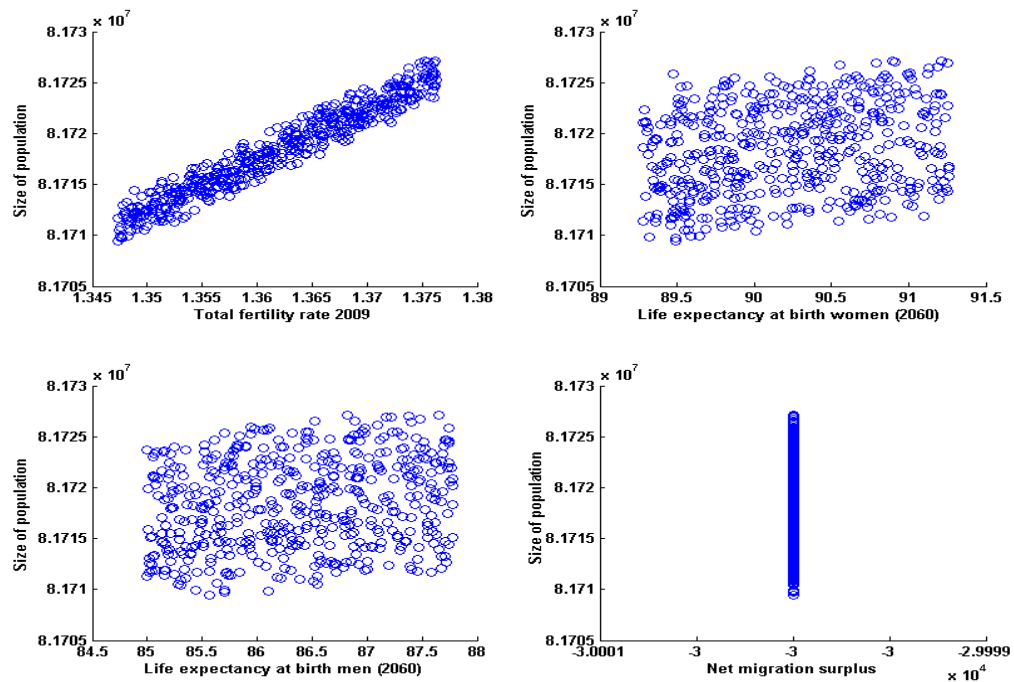
Figure 5 Youth dependency and old age dependency ratio



5.1 Visual sensitivity analysis with scatter-plots

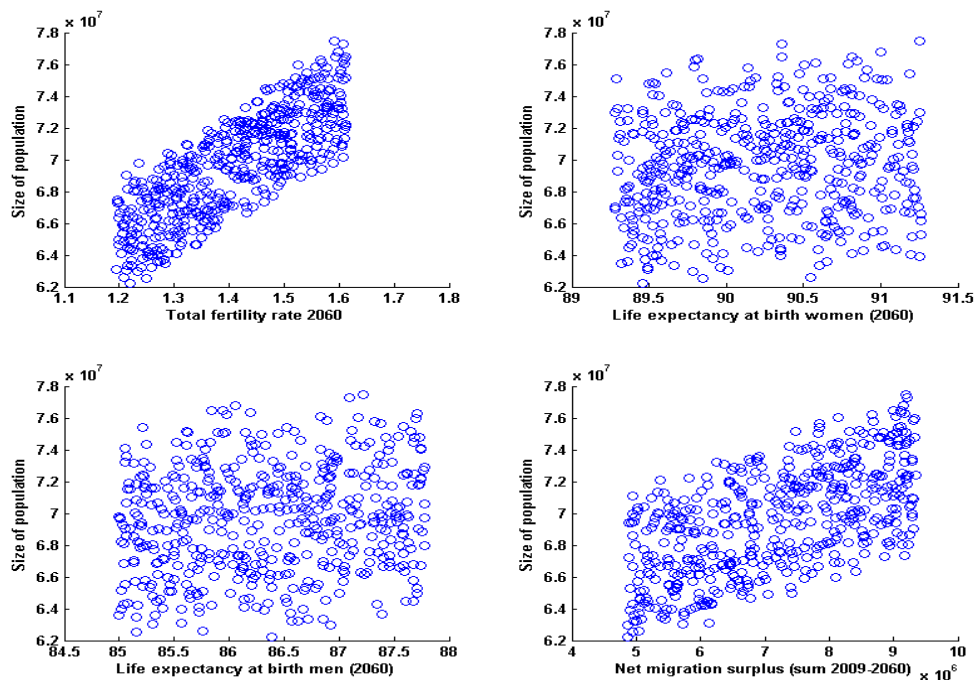
As a first step of the investigation of sensitivities scatter-plots are a proven tool to visualise the relative importance of the model inputs for the model output. The Figures 6-9 show for different time points (i.e. 2009, 2060) the N values obtained for the three model outputs (Size of population, youth dependency ratio, old-age dependency ration) on the y-axis versus the N sample points from the input sample for the model inputs (Total fertility rate, life expectancy females, life expectancy males, net migration surplus) on the x-axis. A scatter-plot with little 'shape', i.e. presenting a rather uniform clouds of points over the range of the input factor on the x-axis, is a sign that the parameter is less influential.

Figure 6 Scatter-plot Size of population vs. model inputs (T=2009)



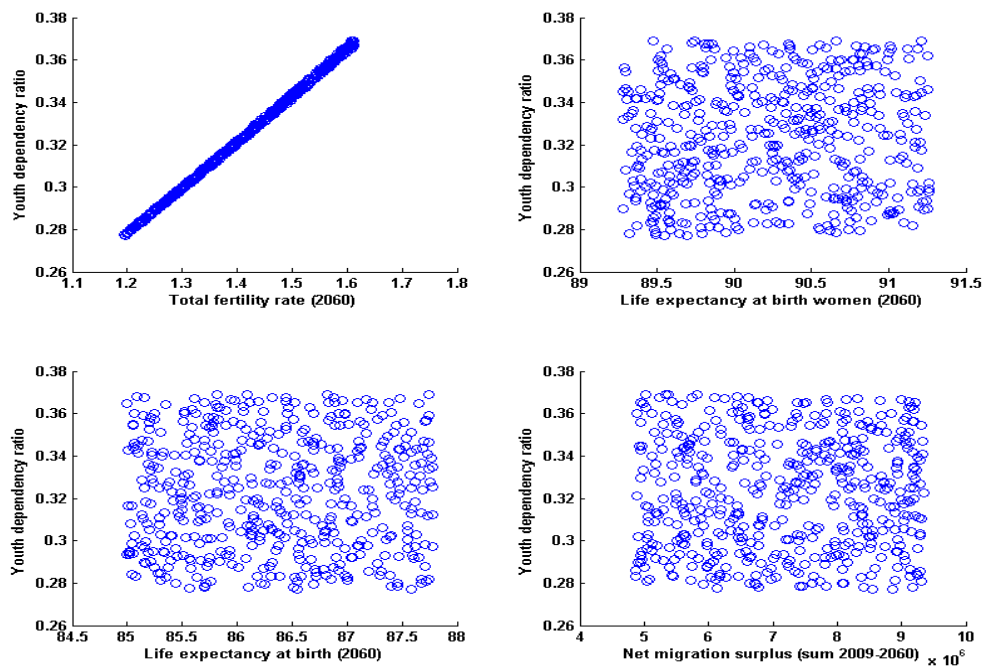
In the first period of the population projection (T=2009) we find a strongly positive correlation of size of population and the total fertility rate. Also a slightly positive relation can be found between size of population and life expectancy. The net migration surplus does not have an impact on the size of population. This is due to the fact that the population projection for 2009 foresees the same net migration surplus of -30.000 persons in the scenarios M1 and M2.

Figure 7 Scatter-plot Size of population vs. model inputs (T=2060)



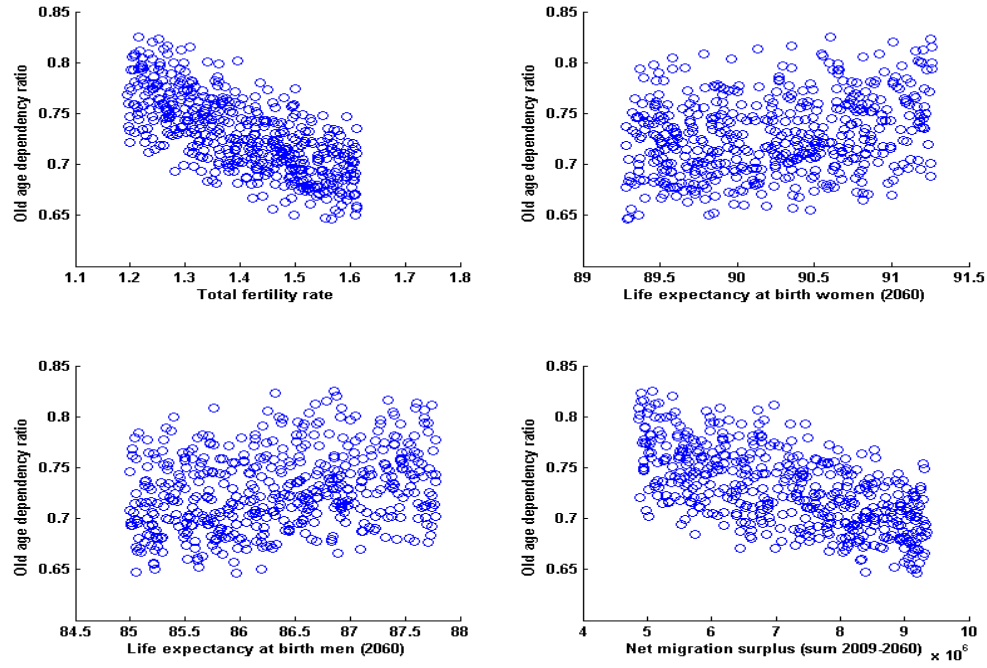
The picture is different if we look on the last year of the projection horizon ($T=2060$). There is still a positive relation between size of population and total fertility rate, but the band has widened. The other model inputs account now for a larger share of the variance in the size of population. If one looks at the relation between size of population and life expectancy, one can hardly detect a trend in the cloud of sample points. For 2060 horizon also the net migration surplus shows a positive relation with respect to the size of population even though to a lower extent in comparison with the total fertility rate.

Figure 8 Scatter-plot Youth dependency ratio vs. model inputs ($T=2060$)



Regarding the youth dependency ratio we have plotted only the scatter-plots referring to the last period of the population projection ($T=2060$). The results over the entire projection horizon (2009-2060) look very similar to those presented in Figure 8. There is only one model input for which the cloud of points indicates a clear 'shape' - that is for the total fertility rate. The uncertainty in the youth dependency ratio is not explained by any of the other model inputs.

Figure 9 Scatter-plot Old age dependency ratio vs. model inputs (T=2060)

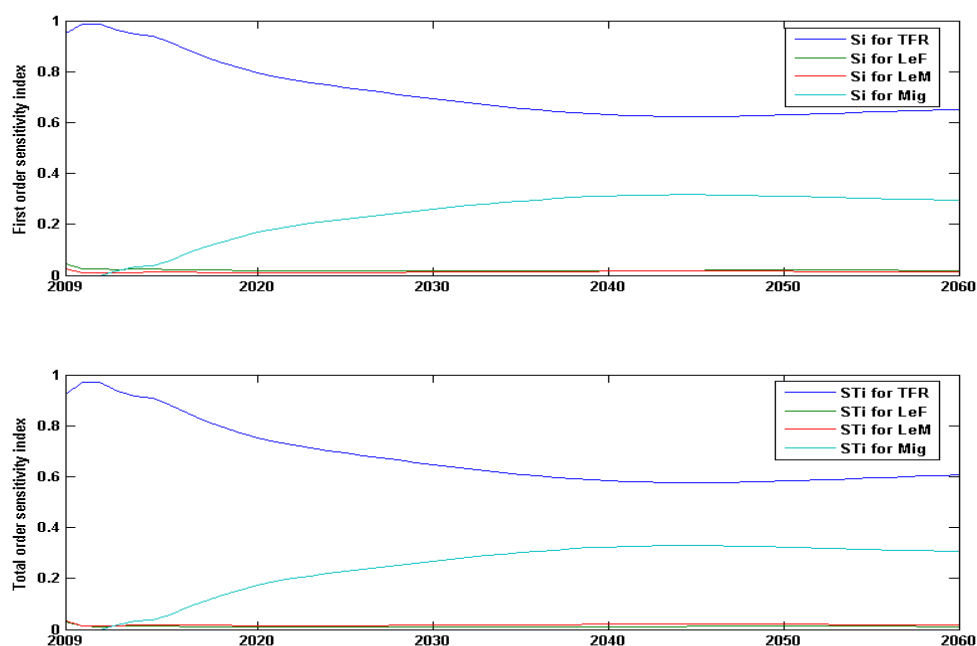


Over the projection horizon the old age dependency ratio shows changing sensitivity indices over the projection horizon. For the projection of the year 2060, we find a negative correlation of old age dependency and total fertility rate, and also for old age dependency ratio and net migration surplus. A slightly positive correlation can be found of old age dependency ratio and life expectancy.

5.2 Estimated sensitivity indices

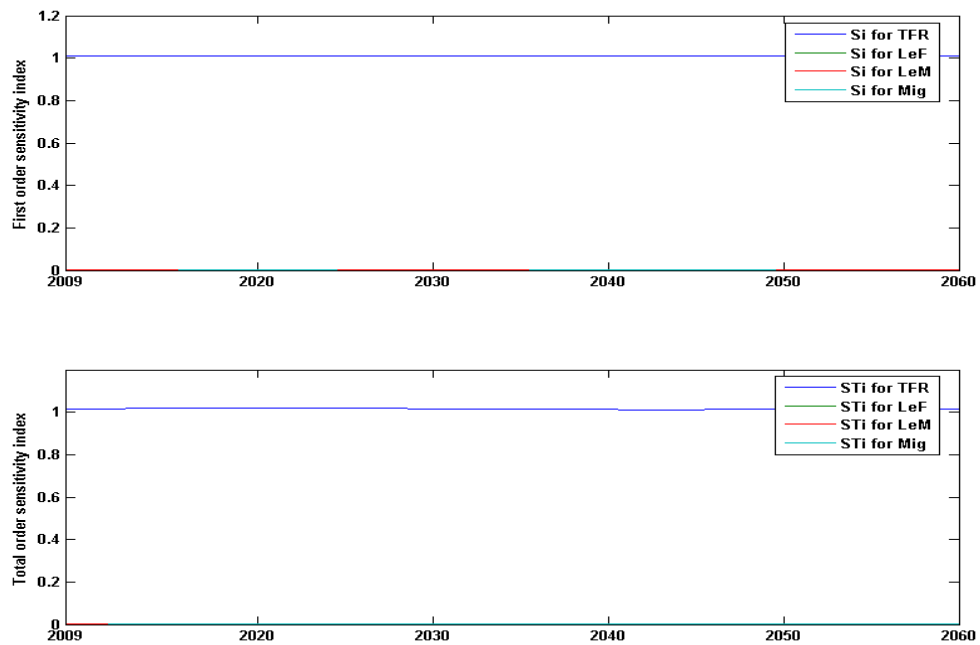
The calculation of the sensitivity indices of first and total order over the entire period of the projection proceeds as described in section 3.2. The total computational costs for the estimation is $N \cdot (k+2)$ model runs. The model runs lead to the following results for the sensitivity estimates:

Figure 10 Sensitivity indices: Size of population



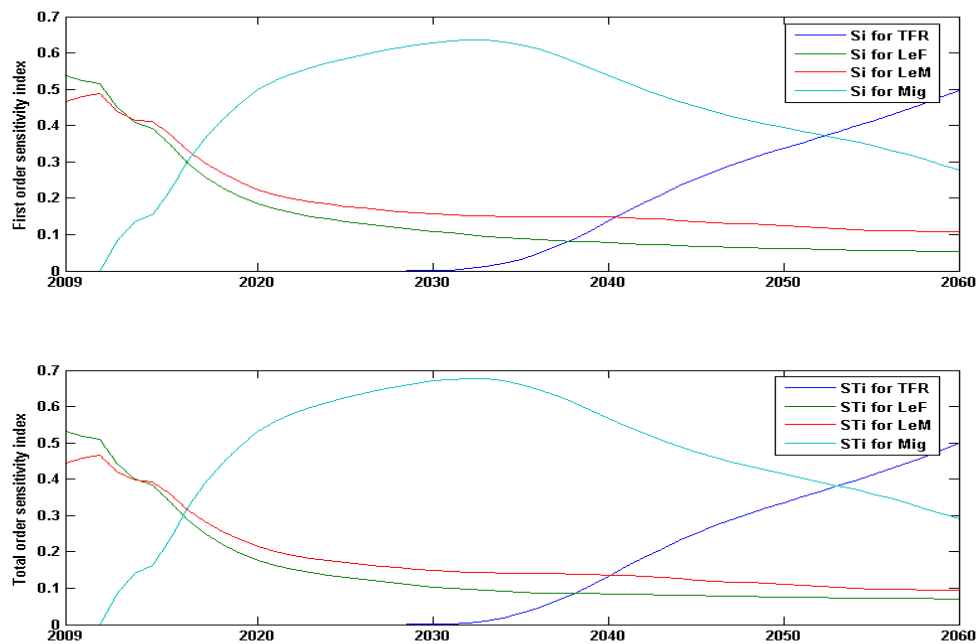
The sensitivity indices for the four model inputs over the projection horizon 2009-2060 indicate that the uncertainty in the assumptions on the total fertility rate drives most of the variance in the size of the population. The first order sensitivity indices for this model input range from 0.66 to 1.05. For the total sensitivity indices we find a range of 0.58 to 0.92. The uncertainty in the size of the population is only fractionally driven by the assumptions on life expectancy. The largest value for the sensitivity of this model output we can identify for the first year of the population projection (2009), although the values remain below 0.03. The sensitivity indices for the net migration surplus increase as the projection horizon extends until 2040 and then remain relatively constant with values around 0.31-0.35.

Figure 11 Sensitivity indices: Youth dependency ratio



The uncertainty in the youth dependency ratio over the entire projection horizon is only affected by the uncertainty in the total fertility rate. The further model inputs such as life expectancy for women and men and migration do not explain any of the variance in the projected youth dependency ratio.

Figure 12 Sensitivity indices: Old age dependency ratio



The third model output investigated is the old age dependency ratio. For this model output the sensitivity indices vary over the projection horizon for all model inputs. For the first years of the population projection the assumptions about life expectancies drive most of the uncertainty. The total fertility rate does not explain any of the variance in the old age dependency ratio before 2028. Starting from this year until the end year of the projection the sensitivity indices are strongly increasing. From 2053 the factor even causes most of the uncertainty in the model output.¹⁰ Over most years of the entire projection horizon, the net migration surplus is the main driver of uncertainty (2014-2052). The highest values of the sensitivity indices for this value is reached in 2033 when around 70% of the variance in the old age dependency ratio can be due to the model parameter. From this year, the relative importance of this model input starts declining and finally reaches 30% in 2060. At the end of the projection horizon (T=2060), 50% of the variance in the projected values is driven by the total fertility rate, the life expectancies explain 22% (9% life expectancy women, 11% life expectancy men) and the net migration surplus accounts for 30% of the variance.

Table 1 First order sensitivity indices

	2009	2020	2030	2040	2050	2060
	<i>Size of population</i>					
Fertility rate	1.05	0.86	0.74	0.67	0.66	0.69
Life expectancy women	0.01	0.00	0.00	0.00	0.00	0.00
Life expectancy men	0.03	0.01	0.01	0.02	0.02	0.01
Net migration surplus	0.00	0.19	0.29	0.35	0.35	0.33
	<i>Youth dependency ratio</i>					
Fertility rate	1.11	1.11	1.11	1.11	1.11	1.11
Life expectancy women	0.00	0.00	0.00	0.00	0.00	0.00
Life expectancy men	0.00	0.00	0.00	0.00	0.00	0.00
Net migration surplus	0.00	0.00	0.00	0.00	0.00	0.00
	<i>Old age dependency ratio</i>					
Fertility rate	0.00	0.00	0.00	0.15	0.38	0.55
Life expectancy women	0.55	0.20	0.12	0.10	0.10	0.09
Life expectancy men	0.49	0.25	0.18	0.16	0.13	0.11
Net migration surplus	0.00	0.58	0.73	0.62	0.45	0.32

Table 2 Total order sensitivity indices

	2009	2020	2030	2040	2050	2060
	<i>Size of population</i>					
Fertility rate	0.92	0.75	0.65	0.58	0.58	0.61
Life expectancy women	0.03	0.01	0.01	0.01	0.01	0.01
Life expectancy men	0.03	0.01	0.02	0.02	0.02	0.02
Net migration surplus	0.00	0.17	0.27	0.32	0.32	0.31
	<i>Youth dependency ratio</i>					
Fertility rate	1.05	1.05	1.05	1.05	1.05	1.05

¹⁰ This is can be easily explained by becoming aware of what is measured by the old age dependency ratio. The first year of the projection horizon for which differences in the total fertility rate matter, is the one in which the newborn of the first year of the projection (T=2009) become part of the working age population (from 20-64 years of age), which is not the case in during the projection covering the first 20 years.

Life expectancy women	0.00	0.00	0.00	0.00	0.00	0.00
Life expectancy men	0.00	0.00	0.00	0.00	0.00	0.00
Net migration surplus	0.00	0.00	0.00	0.00	0.00	0.00
	<i>Old age dependency ratio</i>					
Fertility rate	0.00	0.00	0.00	0.14	0.35	0.51
Life expectancy women	0.55	0.20	0.11	0.09	0.08	0.07
Life expectancy men	0.44	0.23	0.16	0.14	0.11	0.09
Net migration surplus	0.00	0.59	0.73	0.61	0.44	0.31

If we compare the sensitivity indices presented in Tables 1 and 2 we find almost no difference between first and total order sensitivity indices. This indicates that the cohort-component method can be classified as additive model. Interactions between model inputs are not present.

6 Conclusions

In this study I apply variance-based sensitivity analysis in the area of population projections. This allows to decompose the variance in the output of the demographic model into its driving factors and thereby to quantify the role of the model inputs. In the sensitivity analysis setting of the cohort-component method four uncertain model inputs (total fertility rates, life expectancy of women and men, and the net migration surplus) and three model outputs (size of population, youth dependency ratio, old age dependency ratio) were investigated. The range of uncertainty in the model inputs are given by the assumptions made for those model inputs in the 12th population projection for Germany.

Instead of the scenario approach most often employed in population projections, we have used samples within the range of the assumptions taken. By doing so, the number of scenarios increases from 12 to 500 which allow us to apply more sophisticated statistical tools.

The results show that the assumptions on the total fertility rate drive most of the uncertainty in two of the three model outputs (size of population, youth dependency ratio). However, the calculated sensitivity indices vary over the projection horizon (in particular for the old age dependency ratio). The assumed increase in life expectancies plays a minor role as driver for the uncertainty in the size of population except for the first years of the projection for the old age dependency ratio. For the latter model output, over the projection horizon 2014-2052, the net migration surplus contributes the largest share to the uncertainty in the old-age dependency ratio. The uncertainty in the projection horizon for 2053-2060 is mainly driven by the assumptions on the total fertility rate.

In order to increase the knowledge about what is causing the variance in the population projection over time, the application of variance based sensitivity analysis can be extremely useful. A drawback of the approach is, however, that the sensitivity indices presented here are only valid for the assumptions made in a particular test case but cannot be generalised to any population projection. The results for sensitivity indices would differ if another dataset (for instance another country with different population age structure) was used. An advantage of the approach is that instead of scenarios just some very basic framework about the most likely paths for fertility, life

expectancy and migration would be required to run a population projection. Within those bounds, samples can be generated which lead to a much more higher number of projected values for the model outputs, and additionally allow the attribution of uncertainties in the model output to the ones in the model inputs.

7 Acknowledgements

I would like to thank my colleague Mr Stefano Tarantola from the Institute for the Protection and Security of the Citizen, Italy for his support in implementing the cohort component method in a sensitivity analysis setting. Furthermore, I thank Ms Olga Poetzsch from the Federal Statistical Office, Germany for the provision of some more detailed information on the assumptions taken in the 12th population projection for Germany.

References

- Ahlburg, D.A. and Lutz, W. (1998). Introduction: The need to rethink approaches to population forecasts. In: Lutz, W., Vaupel, J.W., and Ahlburg, D.A. (eds.). *Frontiers of Population Forecasting*, supplement to Population and Development Review 24: 1-14.
- Betz, F., Lipps, O. (2006). A Stochastic Population Projection for Germany.
- Cukier, R. I., Fortuin, C.M., Schuler, K.E., Petschek, A.G., and Schiably, J.H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients, I Theory, *The Journal of Chemical Physics* 59: 3873-3878.
- European Commission, Directorate General for Employment and Social Affairs (2008). The Demography Report 2008: Meeting Social Needs in an Ageing Society.
- European Commission and Economic Policy Committee (2009). Sustainability Report 2009.
- European Commission and Economic Policy Committee (2009). The 2009 Ageing report.
- European Commission and Economic Policy Committee (2009). The economic consequences of ageing populations: projections for the EU-27.
- Eurostat (2006). Long-term population projections at national level. Statistics in focus.
- Hajnal, J. (1955). Prospects for Population Forecasts. *Journal of the American Statistical Association* 50 (270): 309-322.
- Homma T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of model output. *Reliability Engineering and System Safety* 52(1): 1-17.
- Jansen, M. J. W. (1999). Analysis of variance designs for model output. *Computer Physics Communication* 117: 35-43.
- Keilman, N., Pham, D.Q., and Hetland, A. (2002). Why population forecasts should be probabilistic – illustrated by the case of Norway. *Demographic Research* 15: 409-454.
- Lee, R.D. (1998). Probabilistic approaches to population forecasting. In: Lutz, W., Vaupel, J.W., and Ahlburg, D.A. (eds). *Frontiers of Population Forecasting*, supplement to Population and Development Review 24: 156-190.
- Leslie, P.H. (1945). On the use of matrices in certain population mathematics, *Biometrika* 33: 183-212.

National Research Council (2000). Beyond Six Billion: Forecasting the World's Population. Panel on Population Projections. In: Bongaarts, J. and Bulatao, R.A. (eds); Committee on Population, Commission on Behavioral and Social Sciences and Education. National Academy Press, Washington, DC.

Notestein, F.W. (1945). Population: The long view. In: T.W. Schultz (ed.). *Food for the World*. Chicago: University of Chicago Press: 36-69.

O'Neill, B. C., Balk, D., Brickman, M., Ezra, M. (2001). A guide to global population projections, *Demographic Research* 4: 203-288.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. and Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications* 181: 259–270.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008). Global Sensitivity Analysis. The Primer. Wiley & Sons.

Sobol', I. M. (1990). Sensitivity estimates for nonlinear mathematical models. *Matematicheskoe Modelirovanie* 2: 112-118. in Russian; translated in English as Sobol', I. M., 1993, Sensitivity analysis for non-linear mathematical models. *Mathematical modeling and Computational Experiment* 1: 407-414.

Statistisches Bundesamt (2006). Germany's Population by 2050. Results of the 11th coordinated population projection.

Statistisches Bundesamt (2009). Bevölkerung Deutschlands bis 2060. 12. koordinierte Bevölkerungsvorausberechnung. In German.

U.S. Census Bureau (1999). Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100.

European Commission

EUR 24355 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: Assessing the uncertainty in population projections – A test based on the 12th coordinated population projection for Germany

Author(s): Dirk Zeitz

Luxembourg: Publications Office of the European Union

2010 – 26 pp. –

EUR – Scientific and Technical Research series – ISSN 1018-5593

ISBN 978-92-79-15709-7

DOI 10.2788/90191

Abstract

In this paper I use variance based sensitivity analysis to investigate the uncertainty in projections of future population arising from the assumptions taken. Instead of for a limited number of scenarios the uncertainty is explored within the range of assumptions for the model parameters through sampling.

To this purpose, the cohort component method and the assumptions taken in 12th coordinated population projection for Germany have been reengineered in a sensitivity analysis setting. This allows attributing the uncertainty in the outputs of the demographic model (e.g. size of the population) to the uncertainty in the model inputs (e.g. total fertility rates).

How to obtain EU publications

Our priced publications are available from EU Bookshop (<http://bookshop.europa.eu>), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

